International Journal of Civil Engineering and Technology (IJCIET)
Volume 8, Issue 3, March 2017, pp. 432–440 Article ID: IJCIET_08_03_044
Available online at http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=8&IType=3
ISSN Print: 0976-6308 and ISSN Online: 0976-6316

CONSTRUCTION OF META CLASSIFIERS FOR ACADEMIC RESEARCH DATA FROM SOCIAL NETWORKS

G. Ayyappan
Research Scholar, Department of Computer Science and Engineering,
Bharath University, Chennai, India

Dr. C. Nalini
Professor, Department of Computer Science and Engineering,
Bharath University, Chennai, India

Dr. A. Kumaravel
Professor, School of Computing, Bharath University, Chennai, India

ABSTRACT

Study of research progress in the academic domain is challenging for research communities and funding agencies. The data recovered from the social networks augment this issue for supporting the results in this direction. Here in this paper we address this issue positively with the help text mining tasks. Classification as one of the major data mining methodologies can be applied effectively for this purpose. The objective of this paper is to check the learning algorithms for classification such examples based on selected dataset for research articles in technical conferences. The main intention in this context is to deal with available data set for high accuracy. For this purpose AdaboostM1, Bagging, Dagging, OrdinalClass Classifiers, Stacking models are built using an open source mining Weka under supervised learning algorithms. It is necessary to reduce the error before constructing the final models and thus the varying the parameters and number of iterations for training is carried out.

Key words: Data mining, Classification, Meta classifiers, Base Classifiers, AdaboostM1, Bagging, Dagging, Ordinal Class Classifiers, Stacking.

Cite this Article: G. Ayyappan, Dr. C. Nalini and Dr. A. Kumaravel, Construction of Meta Classifiers for Academic Research Data from Social Networks. *International Journal of Civil Engineering and Technology*, 8(3), 2017, pp. 432–440.
http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=8&IType=3

1. INTRODUCTION

In this paper we address the problem of academic social network data in research progress prediction based on civil and computer science engineering conferences. we present a novel

meta-feature generation method in the context of meta-learning, which is based on rules that compare the performance of individual base learners in a one-against-one manner. Experimental results are based on a large collection of datasets and show that the proposed new techniques can improve the overall performance of meta-learning for algorithm ranking significantly [1].

Nikita Bhatt et al [12] discussed the different approaches of Meta learning based on dataset characteristics provides a system that automatically provides ranking of the classifiers by considering different characteristics of datasets and different characteristics of classifiers after the generation of the Meta Knowledge Base, Ranking is provided based on Adjusted Ration of Ratio (ARR) or accuracy or time that helps non-experts in algorithm selection task.

Artur Ferreira et al [5] presented an overview of boosting algorithms to build ensembles of classifiers. The basic boosting technique and its variants are addressed and compared for supervised learning. The extension of these techniques for semi-supervised learning is also addressed. For face detection, boosting algorithms have been the most effective of all those developed so far, achieving the best results.

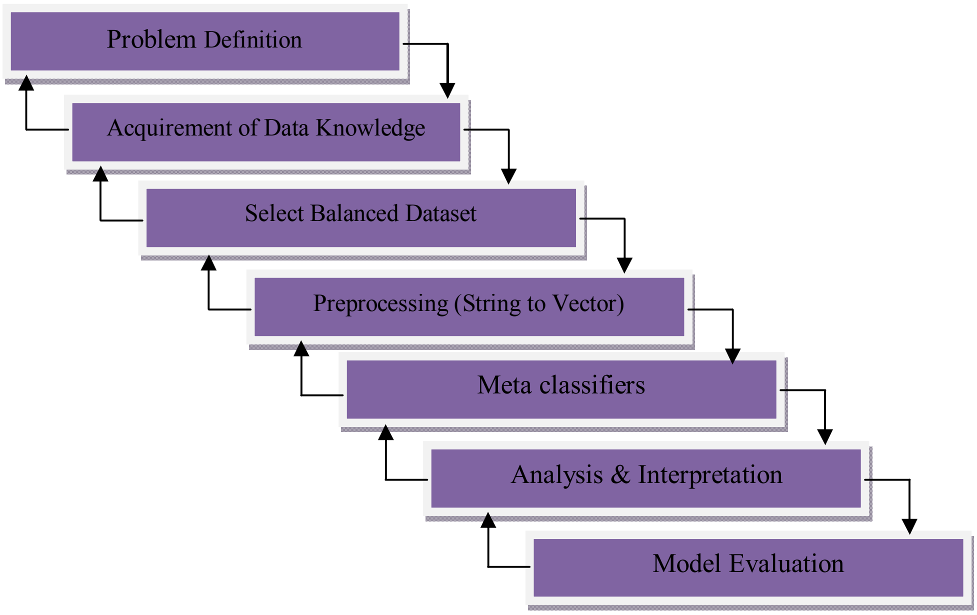

Figure 1 Data Mining Iteration Steps

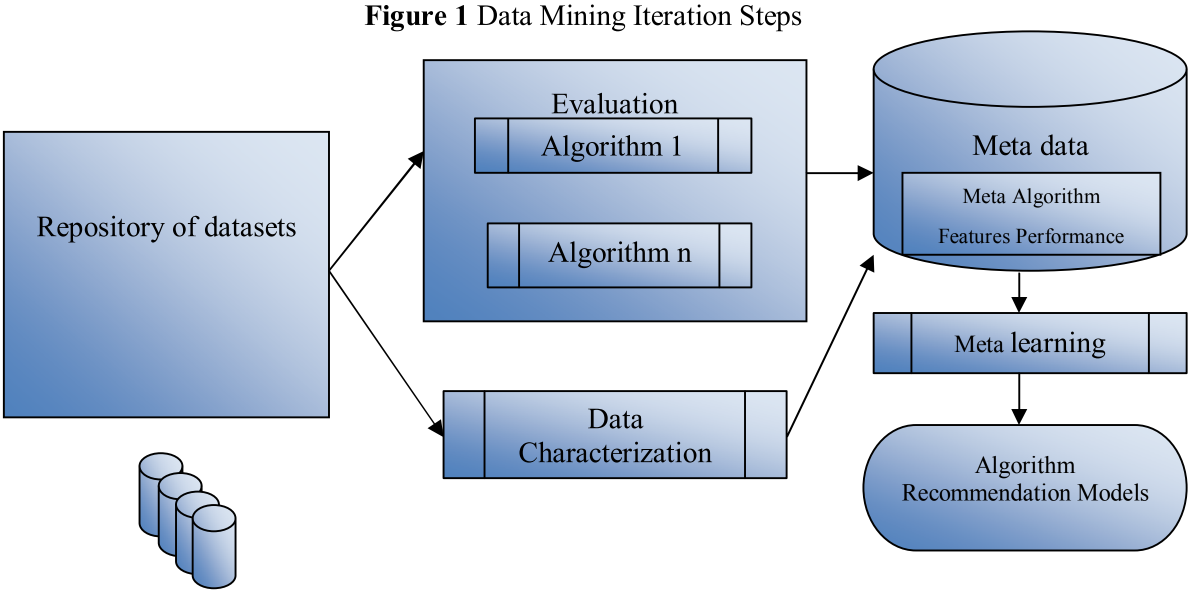

Figure 2 System for constructing meta-classifier

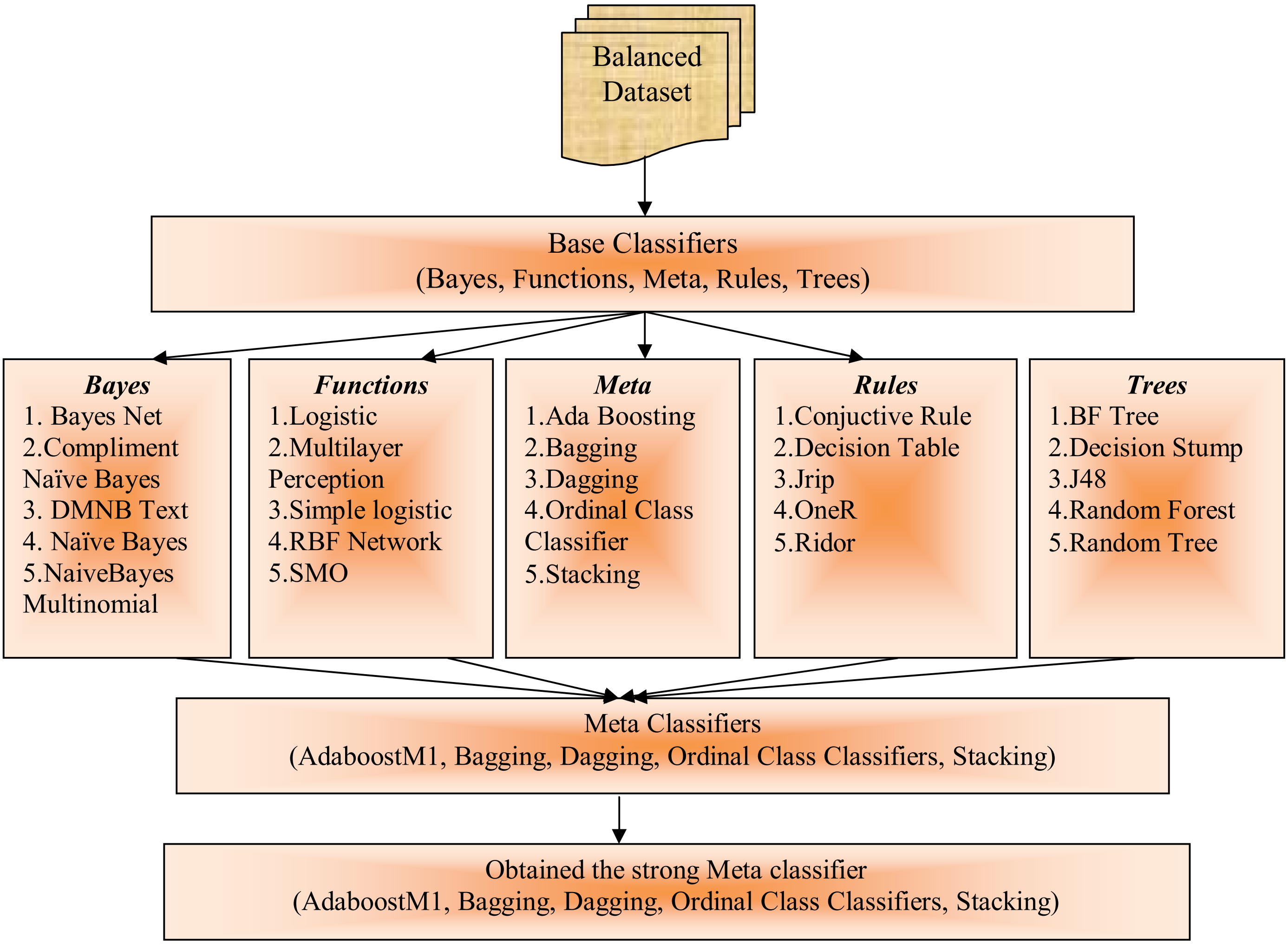

Figure 3 Flow to obtain a Strong Meta classifier

2. META CLASSIFIERS

Meta Classifier has showed spectacular success in reducing classification error from learned classifiers. These techniques develop a classifier in the form of a committee of classifiers. The committee members are applied to a classification task and their individual outputs combined to create a single classification. Meta learning approaches like AdaBoostM1, Bagging, Dagging, Ordinal Class Classifiers, and Stacking [2,3,10,11]Parameter Selection have received extensive attention. They are the recent methods for improving the predictive power of classifier learning systems.

Table 1 Functions of classifiers

Meta Classifier name	Category	Functions
Meta	*Adaboost M1*	Class for boosting a nominal class classifier using the Adaboost M1 method.
	Bagging	Bag a classifier; works for regression too
	Dagging	It creates a number of disjoint, stratified folds out of the data and feeds each chunk of data to a copy of the supplied base classifier.
	OrdinalClassClassifier	Meta classifier that allows standard classification algorithms to be applied to ordinal class problems.
	Stacking	Combines several classifiers using the stacking method.
Base	*BayesNet*	Numeric estimator precision values are chosen based on analysis of the training data
	Compliment NaiveBayes	Class for building and using a Complement class Naive Bayes classifier.
	DMNB Text	Class for building and using a Discriminative Multinomial Naive Bayes classifier.
	NaiveBayes	Class for a Naive Bayes classifier using estimator classes.
	NaiveBayes Multinomial	Class for building and using a multinomial Naive Bayes classifier.
	Logistic	Class for building and using a multinomial logistic regression model with a ridge estimator.
	MultilayerPerception	A Classifier that uses back propagation to classify instances. This network can be built by hand, created by an algorithm or both.
	Simplelogistic	Classifier for building linear logistic regression models.
	RBFNetwork	Class that implements a normalized Gaussian radial basis function network.
	SMO	Implements John Platt's sequential minimal optimization algorithm for training a support vector classifier.
	Adaboost M1	Class for boosting a nominal class classifier using the Adaboost M1 method.
	Bagging	Bag a classifier; works for regression too
	Dagging	It creates a number of disjoint, stratified folds out of the data and feeds each chunk of data to a copy of the supplied base classifier.
	OrdinalClassClassifier	Meta classifier that allows standard classification algorithms to be applied to ordinal class problems.
	Stacking	Combines several classifiers using the stacking method.
	ConjuctiveRule	This class implements a single conjunctive rule learner that can predict for numeric and nominal class labels.
	Decision Table	Class for building and using a simple decision table majority classifier.
	JRip	This class implements a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER),
	OneR	Class for building and using a 1R classifier; in other words, uses the minimum-error attribute for prediction, discretizing numeric attributes.
	Ridor	An implementation of a RIpple-DOwn Rule learner.
	BFTree	Class for building a best-first decision tree classifier.
	DecisionStump	Usually used in conjunction with a boosting algorithm.
	J48	For generating a pruned or unpruned C4.5 decision tree.
	RandomForest	Class for constructing a forest of random trees.
	RandomTree	Class for constructing a tree that considers K randomly chosen attributes at each node.

3. EXPERIMENTAL ANALYSIS

In this section, we test the implementation efficiency of various methods and compare with whole dataset and the selected attributes. Weka tool is used to construct classification models.

3.1. Dataset

The datasets for these experiments are from [7].The original data format has been slightly modified and extended in order to get relational format.

3.1.1. Dataset Information

The database of academic social network this dataset describes a set for selected attributes for Best first search method in the range as shown in the table 1. The output is categorized into large, medium, small. The output class is denoting the possible category of infection affected. Number of Instances in this database is 6000.

Table 2 List of Attribute and their Data Type

Attribute Name	Data Type	Minimum	Maximum	Mean	Standard Deviation
2010	Numeric	0	1	0.058	0.233
201th	Numeric	0	1	0.03	0.171
Artificial	Numeric	0	1	0.029	0.168
Conference	Numeric	0	1	0.214	0.41
Microsoft	Numeric	0	1	0.005	0.067
University	Numeric	0	1	0.208	0.406
accurate	Numeric	0	1	0.015	0.121
recognition	Numeric	0	1	0.028	0.164
systems	Numeric	0	1	0.201	0.401
Algorithm	Numeric	0	1	0.02	0.139
Approximation	Numeric	0	1	0.007	0.083
Architecture	Numeric	0	1	0.009	0.094
Object - Oriented	Numeric	0	1	0.007	0.083
Soccer	Numeric	0	1	0.007	0.085
Operations	Numeric	0	1	0.004	0.062
Turku	Numeric	0	1	0.007	0.036
Infection Class	Nominal	No of Classes 3			

4. METHODOLOGY

The first step of our analysis was to reduce the high data dimensionality. For this purpose we used Weka tool [4,6,8,9,13] for attribute selection based on various search methods16 made in the attribute space as shown in table 1. We used factors which are selected after preprocessing as new predictors.

4.1. Method Description

Here we use three meta data classifications with different iterations in Ada Boost is decision stump and getting 57.2464 accuracy in my represented iterations so skipped this and Bagging is Rep tree classifier with maximum of 79.7101commonly for three iterations ,in Logit Boost is Decision stump classifier with maximum of 76.8116.

4.2. Ada Boost

This is meant for boosting a nominal class classifier method. Only nominal class problems can be tackled. Often dramatically improves performance, but sometimes over fits.

4.3. Bagging

This is meant for bagging a classifier to reduce variance. It can do classification and regression depending on the base learner. Logit Boost This is meant for performing additive logistic regression. This model enables to classify the dataset using regression method as base classifier and can be applied for more than binary class problems. The following tables show the performance of the above methods in Weka which is a java implementation.

Table 3 Various Classifiers accuracies

Base Classifiers \ Meta Classifiers		AdaBoost M1	Bagging	Dagging	Ordinal Class Classifiers	Stacking
Bayes	Bayes Net	72.18	72.18	69.3	71.65	33.33
	Compliment Naïve Bayes	52.96	52.95	53.02	52.41	33.33
	DMNB Text	69.86	70.26	33.9	69.46	33.33
	NaiveBayes	69.98	70.1	70.21	69.61	33.33
	NaiveBayes Multinomial	49.56	56	49.7	52.15	33.33
Functions	Logistic	72.13	72.05	71.88	71.73	33.33
	Multilayer Perception	71.98	58.15	42.81	71.91	33.33
	Simplelogistic	72.11	72.13	72.01	45.56	33.33
	RBFNetwork	42.85	65.15	71.5	70.41	33.33
	SMO	71.75	71.75	62.41	71.96	33.33
Meta	Adaboost M1	51.38	51.38	54.21	68.11	33.33
	Bagging	71.65	72.1	71.5	71.88	33.33
	Dagging	70.66	71.38	67.28	70.05	33.33
	Ordinal Class Classifier	71.03	70.73	67.16	70.48	33.33
	Stacking	33.33	33.33	33.33	33.33	33.33
Rules	Conjuctive Rule	51.38	51.38	60.43	49.13	33.33
	Decision Table	71.65	72.1	70.41	70.9	33.33
	JRip	70.66	71.06	70.28	70.16	33.33
	OneR	51.38	51.38	51.38	49.23	33.33
	Ridor	69.86	69.82	66.91	68.46	33.33
Trees	BFTree	72.03	72.08	71.83	72.06	33.33
	DecisionStump	51.38	51.38	61.53	66.05	33.33
	J48	71.73	71.73	68.41	70.48	33.33
	RandomForest	72.06	72.08	72.1	72.05	33.33
	RandomTree	72.12	72.1	72.06	72.13	33.33

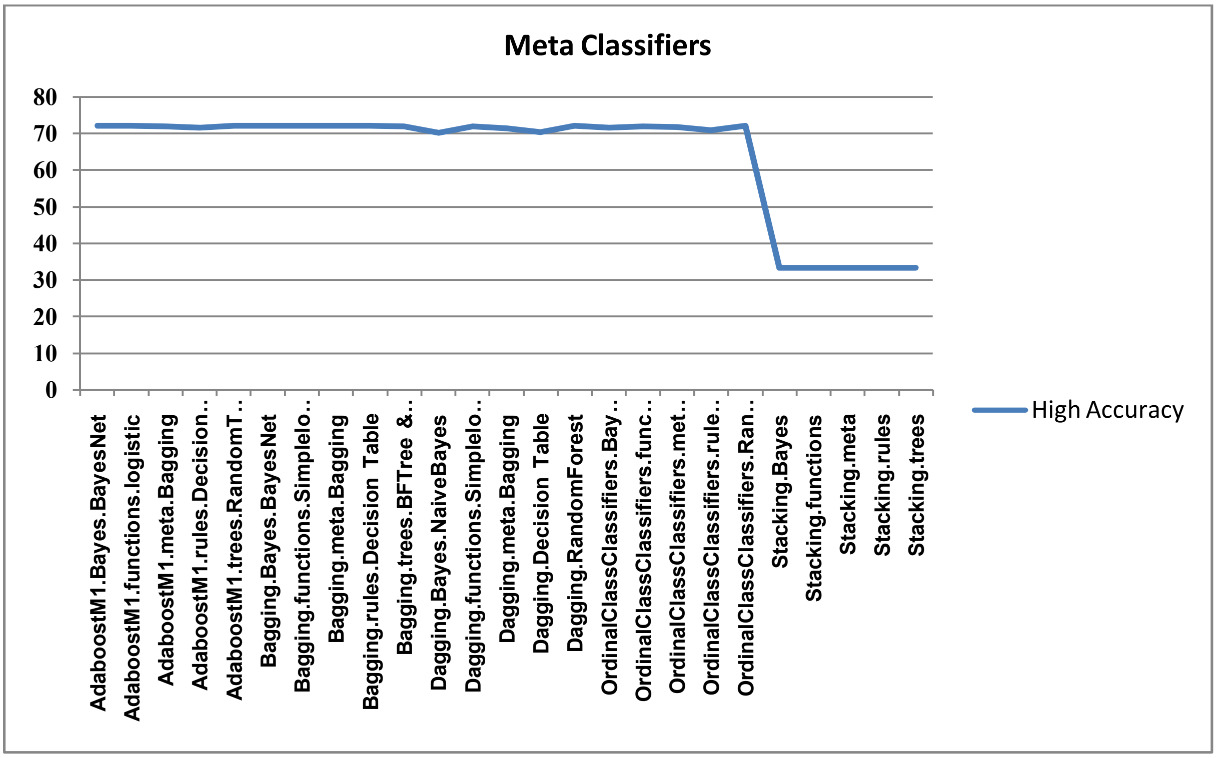

Figure 5 Comparison of meta classifier algorithms for accuracy

5. RESULTS

The Best Meta classifier seen from above tables happens to be AdaboostM1 & Bagging and the parameter values with the BayesNet as base classifier. Then we get best meta classifiers both are same accuracy. So As per ROC the Bagging with BayesNet is produce best accuracy. So we recommended best meta classifiers is bagging with the parameter BayesNetas base classifier.

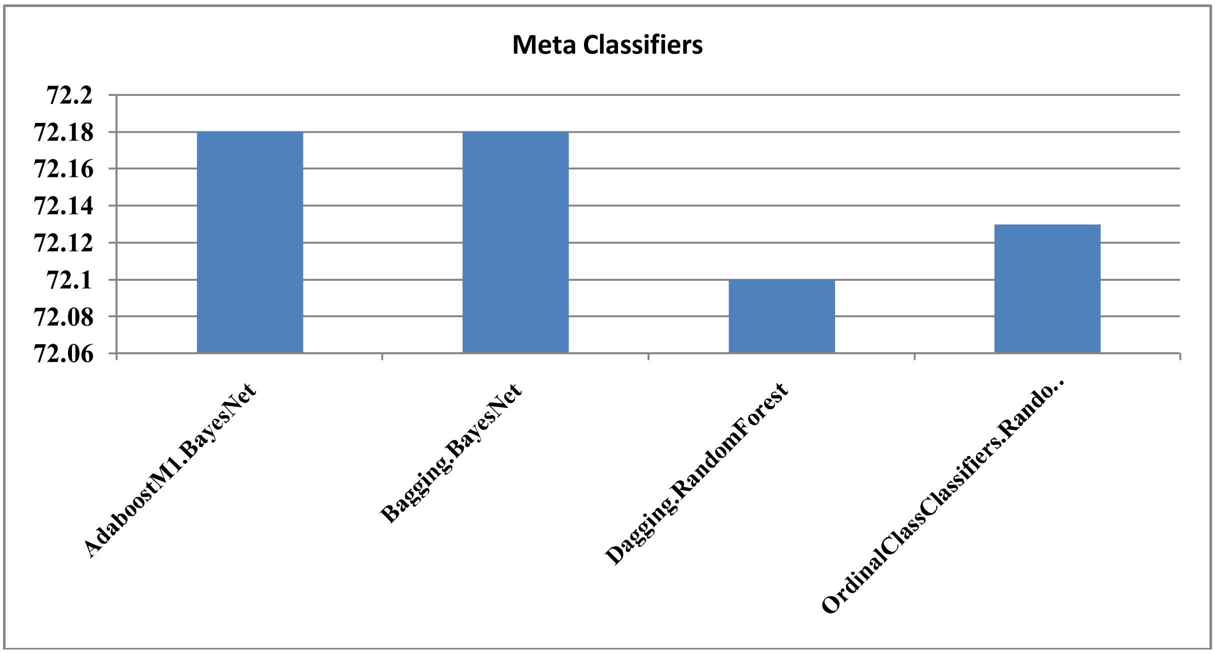

Figure 6 Comparison of Meta classifier algorithms for accuracy

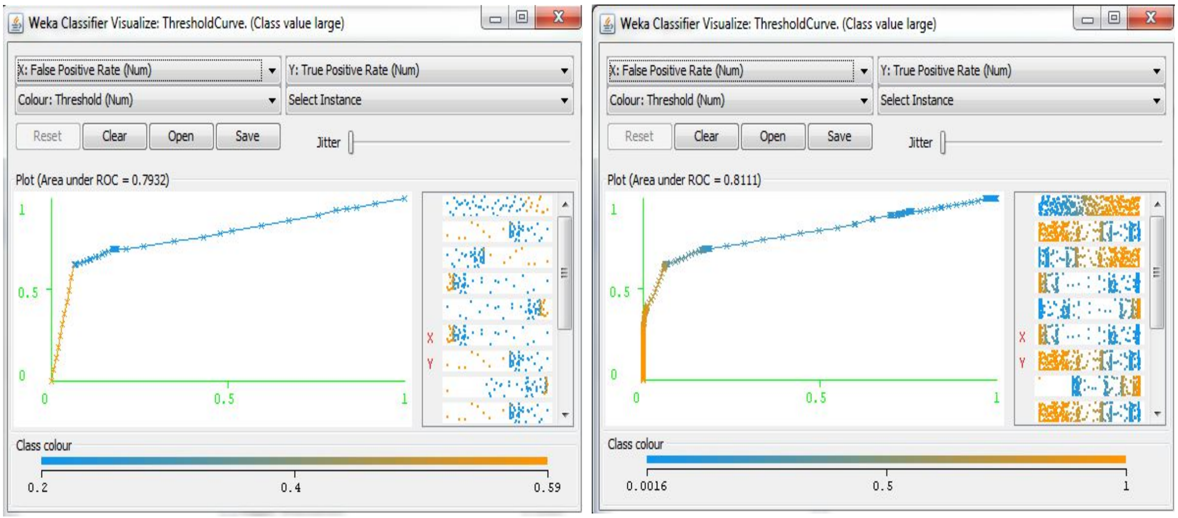

Figure 7 Comparison of AdasBoostM1 and Bagging in dataset named as large

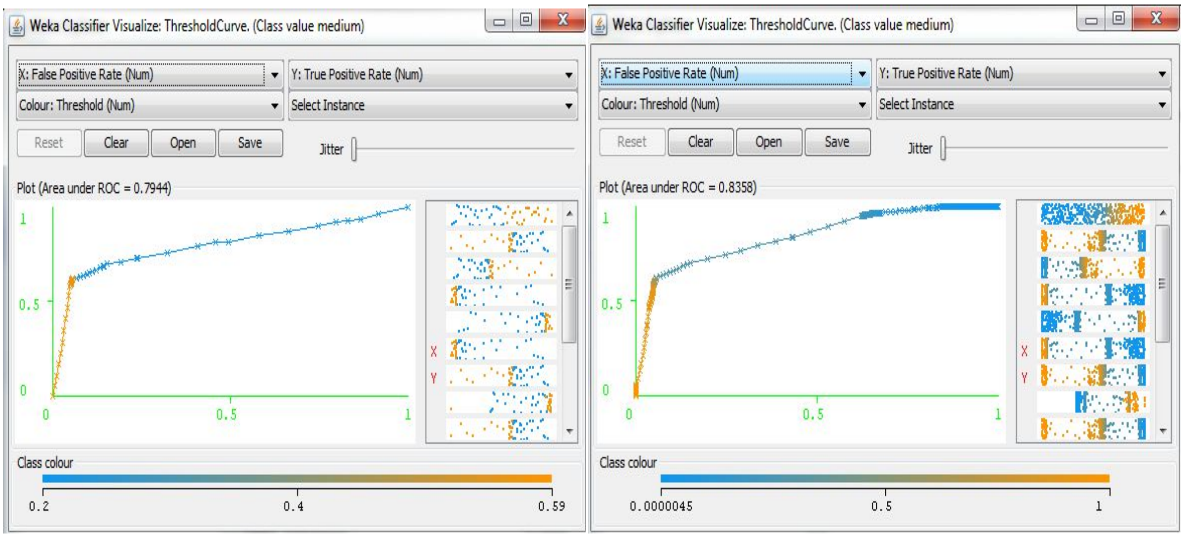

Figure 8 Comparison of AdasBoostM1 and Bagging in dataset named as medium

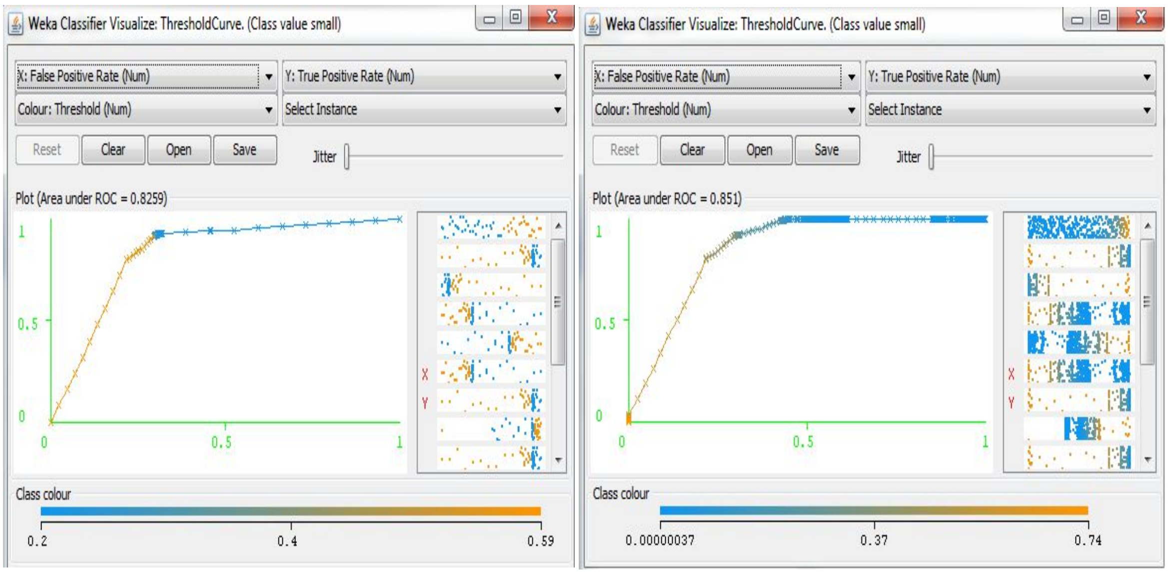

Figure 9 Comparison of AdasBoostM1 and Bagging in dataset named as small

6. CONCLUSION

The above results improve the previously obtained accuracies and this study will help to formulate better schemes for preventing infections and enhancing the yields. However the size of the dataset is not large, future research can accommodate with either large dataset or aggregating small data set into bigger size. I contribute my research works based on academic social network predictions for society of civil engineering and computer science engineering.

REFERENCES

[1] Quan Sun, Pfahringer, “Pairwise meta-rules for better meta-learning-based algorithm ranking Machine learning”, Springer US, Machine Learning, 93(1):141-161, 2013.

[2] Ankit Desai and P M Jadav. Article: An Empirical Evaluation of Ada Boost Extensions for Cost-Sensitive Classification. International Journal of Computer Applications 44(13):34-41, April 2012

[3] http://weka.sourceforge.net/doc.dev/ [online jan’2017]

[4] http://www.ijpbs.net/cms/php/upload/2938_pdf.pdf [online Jan’ 2017]

[5] https://fenix.tecnico.ulisboa.pt/downloadFile/282093452003810/Boosting%20-%20Ferreira%20and%20Figueiredo%202013.pdf [online Jan’ 2017]

[6] Tao wang, Zhenxing Qin, Zhi Jin and Shichao Zhang , "Handling over fitting in test cost-sensitive decision tree learning by feature selection, smoothing and pruning", The journal of systems and software, 2010.

[7] https://aminer.org/data [online December’2016]

[8] Abdullah Wahbeh H, Mohammed Al-Kabi., “Comparative Assessment of the Performance of Three WEKA Text Classifiers Applied to Arabic Text”, Vol. 21, No. 1, pp. 15- 28, 2012.

[9] Books: Data mining - Concept and Techniques by Han & Kamber. Data mining: concepts and techniques. The Morgan Kaufmann series in data management systems, ISBN-1558609016, 9781558609013, publisher morgan, 2006.

[10] Shaidah Jusoh, Hejab Alfawareh M., “Techniques, Applications and Challenging Issues in Text Mining”, Vol. 9, Issue 6, No 2, November2012.

[11] Shilpa DhanjibhaiSerasiya, Neeraj Chaudhary., “Simulation of Various Classifications results using WEKA”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-1, Issue-3, August 2012.

[12] Nikita Bhatt, Amit Thakkar, Amit Ganatra., “A Survey & Current Research Challenges in Meta Learning Approaches based on Dataset Characteristics”, Volume-2, Issue-1, March 2012.

[13] G. Ayyappan, Dr. C. Nalini and Dr. A. Kumaravel, Efficient Mining for Social Networks Using Information Gain Ratio Based on Academic Dataset. *International Journal of Civil Engineering and Technology,* 8(1), 2017, pp. 936–942.

[14] Malpani Radhika S and Dr. Sulochana Sonkamble, A Data Mining Approach to Avoid Potential Biases. *International Journal of Computer Engineering and Technology*, 6 (7), 2015, pp. 27-34.